APPENDIX

Let $\boldsymbol{w}$ be a 3D point in world coordinates, $\boldsymbol{c}$ be the location of the camera in that coordinate system, R be a rotation matrix representing the orientation of the camera (see Figure A1). A local Cartesian camera coordinate system is specified by axes X-Y-Z, where the Z axis lies along the central viewing ray of the camera and X and Y are perpendicular to that ray. We represent points in the image using film coordinates (u , v) on a film plane displaced from the optic center by a distance f, the focal length. Note that film coordinates are related to image pixel coordinates through an affine transformation that is a function of linear camera intrinsic parameters, and also perhaps by an additional nonlinear transformation representing lens distortion. Working in film coordinates amounts to assuming a calibrated camera, for which intrinsic parameters and lens distortion are known and invertible.

In homogeneous coordinates, we have the projection equation

$$[u\ v\ f]' \sim R\,(w - c)$$

relating film coordinates to 3D world coordinates, where ~ denotes equality up to a scale factor and ' denotes a vector or matrix transpose.

A1 : Camera rotation induces a 2D homography in film coordinates.

Let $[u_1\ v_1\ f]'$ be film coordinates for the image taken with orientation R_1, and let $[u_2\ v_2\ f]'$ be film coordinates for the image taken with orientation R_2. Then

$$[u_1\ v_1\ f]' \sim R_1\,(w - c)$$

and

$$[u_2\ v_2\ f]' \sim R_2\,(w - c) \sim R_2 R_1'\,[u_1\ v_1\ f]'$$

Therefore the change in camera orientation induces a 2D homography $R_2 R_1'$ on points in the film plane, and the image can be transformed with no knowledge of scene structure.

A2 : Change in zoom induces an isotropic scaling of film coordinates.

Let $[u_1, v_1, f_1]'$ be film coordinates for the image taken with focal length f_1, and let $[u_2, v_2, f_2]'$ be film coordinates for the image taken with focal length f_2. Then

$$[u_2\ v_2\ f_2]' = (f_2 / f_1)\,[u_1\ v_1\ f_1]'$$

and thus change in focal length induces an isotropic scaling of film coordinates by the magnification factor f_2/f_1 .

A3 : Translation Tz is approximately an isotropic scaling of film coordinates

Let $[x,y,z]' = R\ (w - c)$ be a 3D world point p represented within the camera coordinate system, and let $[u_1, v_1, f]' = [y/z, x/z, f]$ be its film coordinates. If the camera is translated a small distance Tz along the camera's Z axis, then the new coordinates of p are $[x,y,z\text{-}Tz]'$, which map to film coordinates $[u_2, v_2, f]' = [y/(z\text{-}Tz), x/(z\text{-}Tz), f)$, so that

$$[u_2\ v_2] = (z/(z\text{-}Tz))\ [u_1, v_1]$$

and the point is transformed by an isotropic scaling of film coordinates. Although this is strictly true only for this point, and other points at scene depth z, if the object we are looking at is "shallow" with respect to distance from the camera (variation in z coordinates is small with respect to magnitude of z), then this formula is approximately true for all other points on the object as well. Therefore, translation by a small amount Tz towards or away from a distant object is approximated by an isotropic scaling of film coordinates.

A4 : Translation Tx and Ty is approximately a 2D image translation

Let $[x,y,z]' = R\ (w - c)$ be a 3D world point p represented within the camera coordinate system, and let $[u_1, v_1, f]' = [y/z, x/z, f]$ be its film coordinates. If the camera is translated a small distance Tx and Ty along the camera's X and Y axes, then the new coordinates of p are $[x\text{-}Tx,y\text{-}Ty,z]'$, which map to film coordinates $[u_2, v_2, f]' = [(y\text{-}Ty)/z), (x\text{-}Tx)/z, f)$, so that

$$[u_2\ v_2] = [u_1 - Ty/z,\ v_1 - Tx/z]$$

and the point is transformed by a 2D translation of film coordinates. Although this is strictly true only for this point, and other points at scene depth z, if the object we are looking at is "shallow" with respect to distance from the camera (variation in z coordinates is small with respect to magnitude of z), then this formula is approximately true for all other points on the object as well. Therefore, translation by a small amount Tx and Ty perpendicular to the central viewing axis of the camera (the Z axis) can approximated for distant objects by a 2D translation in film coordinates.

A5 : Composition of transformations

The separate transformations for changing orientation and zoom (or translation Tz) can be combined as

$$[u_2\ v_2\ f_2]' \sim (f_2 / f_1)\ R_2 R_1'\ [u_1\ v_1\ f_1]'$$

A6 : Local parameterization of changing orientation and translation

Given an image, we do not necessarily know it's absolute orientation R_1 with respect to the world, nor the absolute orientation R_2 of the desired image. We thus write the homography for changing camera orientation in terms of the current, local camera coordinate. Note that 2D homography $R_2 R_1'$ has the form of a 3D rotation matrix R (up to a scale factor). We can therefore specify the homography R in local camera coordinates as a rotation by angle θ around axis $a = [a\ b\ c]'$. By Rodrigues formula, we can write R as

$$R = I + \sin\theta\, A + (1 - \cos\theta)\, A^2$$

where I is the identity matrix and A is the skew-symmetric matrix

$$A = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}$$

Note, for sufficiently small rotation angles θ, we can use the small angle approximations $\sin\theta \approx \theta$ and $\cos\theta \approx 1$ to write

$$R \approx I + \theta A = \begin{bmatrix} 1 & -c\theta & b\theta \\ c\theta & 1 & -a\theta \\ -b\theta & a\theta & 1 \end{bmatrix}$$

Comparing to the small angle approximation formula for a rotation matrix in terms of Euler angles roll, pitch and yaw about the camera x, y and z axes, respectively (see Figure A1), we see that for small angles: yaw $\approx \theta$a, pitch $\approx \theta$b and roll $\approx \theta$c.

Similar to the case of changing camera orientation, we do not necessarily know the focal length f_1 of a given image, nor the focal length f_2 we would like it to have after a simulated camera zoom. Instead, we parameterize a zoom transformation in terms of a desired magnification factor $m = f_2 / f_1$.

In the case of translation by Tz along the central viewing ray, we do not know the distance z to object points, or the distance Tz of the translation, but we likewise can specify the isotropic scaling induced by the translation as a magnification factor m. Furthermore, translation by Tx and Ty perpendicular to the viewing ray can be specified locally as 2D translation parameters du and dv, which are functions of the unknown object distance z and magnitude of translation components Tx and Ty.

A7 : Specifying yaw and pitch correction by selecting one point correspondence.

Here we show that correction of camera yaw and pitch can be specified by selecting one point correspondence (Figure A2). Point (u,v) is selected, and we wish to determine the camera orientation correction that brings this point to the center of the image (u=0,v=0). That is, we wish to compute the transformation that rotates ray [u v f]' into alignment with ray [0 0 f]'. The unit vector axis of this rotation is

$$[a\ b\ c] = [u\ v\ f]' \times [0\ 0\ f]' = [v\ {-u}\ 0]' / \mathrm{sqrt}(u^2+v^2)$$

and the clockwise angle of rotation is

$$\theta = \mathrm{atan}(\mathrm{sqrt}(u^2+v^2) / f)$$

from which we can write the 2D homography R using Rodrigues formula, above. By construction, the third component of the rotation axis is always zero, i.e. c=0, so that for small angles this is a correction in yaw and pitch alone, with no roll angle correction.

A8 : Efficient approximation for implementing yaw and pitch correction

A great implementation speedup for yaw and pitch correction is possible in the case that we have both small rotation angle and a large focal length. In this case, θ = (sqrt(u^2+v^2) / f) and we can use $R \approx I + \theta A$. After some manipulation we find that the orientation correction homography

$$[u_2\ v_2\ f]' \sim R\,[u_1\ v_1\ f]'$$

reduces to

$$u_2 \approx u_1 - u$$
$$v_2 \approx v_1 - v$$

which is simply the image shift (translation) that takes POI (u,v) to the center of the image.

A9 : Specifying yaw, pitch, roll and Tz by selecting two point correspondences.

Here we show that correction of camera yaw, pitch, roll and Tz can be specified by selecting two point correspondences (Figure A3). As previously (case A7) we select one point as the POI, and determine the 2D homography R that brings this point to the image center (u=0,v=0). In addition to the POI, we also choose a second point, called V1 to denote "vertical unit", which will be mapped one "unit" vertically above the image center. The definition of what distance μ in pixels corresponds to 1 unit is specified by the user (e.g. 1 unit $= \mu = 100$ pixels).

Assumed point V1 is mapped to film coordinate (s,t) by homography R. The coordinate system must be rotated and isotropically scaled so that point (s,t) maps to point $(0,\mu)$ which is located vertically above the image center at distance μ. It can easily be shown that this transformation can be written as

$$S = \begin{bmatrix} \frac{t\mu}{s^2+t^2} & \frac{-s\mu}{s^2+t^2} & 0 \\ \frac{s\mu}{s^2+t^2} & \frac{t\mu}{s^2+t^2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Roll and Tz correction S can be composed with pitch and yaw correction R to yield a single 2D homography $H = S\,R$ that performs pitch, yaw, roll and Tz correction.

A10 : Efficient approximation for changing yaw, pitch, roll and Tz

When implementing for small angles and large focal lengths, the 2D homography H above can be simplified to a translation, rotation and isotropic scale, i.e. a single similarity transformation, leading to a great computational savings. The derivation is similar to the case A8, above.

A11 : Specifying yaw, pitch, roll , Tx, Ty and Tz by selecting three point correspondences.

Here we show that full correction of camera yaw, pitch, roll, Tx, Ty, and Tz can be specified by selecting three point correspondences (Figure A4). As previously (case A9), we select two points to be the POI and V1, and compute the homography that corresponds to correcting yaw, pitch, roll and Tz by bringing the POI to the image center (u=0,v=0) and V1 to point $(0,\mu)$ one vertical unit above the center. Now select a third point C0 = (p,q) in this corrected image. We then want to apply the 2D translation that maps the origin (0,0) to point C0=(p,q), namely

$$u2 = u + p$$
$$v2 = v + q$$

is the correction for Tx and Ty. Let this transformation be written as matrix T. Then the Tx and Ty correction can be composed with the previous pitch, yaw, roll and Tz corrections (R and S) to form a single 2D homography $H = T\,S\,R$ that performs yaw, pitch, roll, Tx, Ty, and Tz.

A12 : Efficient approximation for changing yaw, pitch, roll, Tx, Ty and Tz

When implementing for small angles and large focal lengths, the 2D homography H above can be simplified to a translation, rotation and isotropic scale, i.e. a single similarity transformation, leading to a great computational savings. The derivation is similar to cases A8 and A10, above.

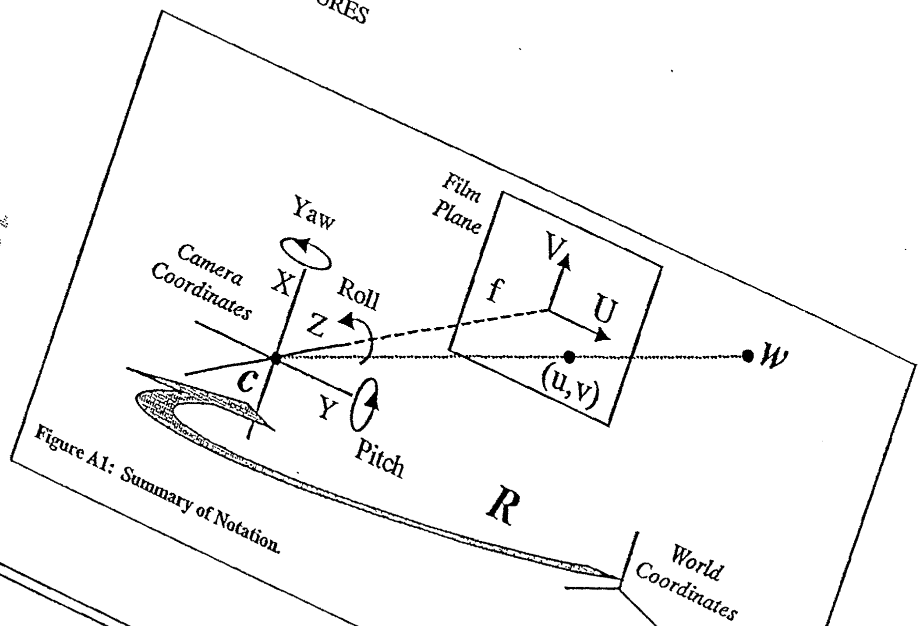

Figure A1: Summary of Notation.

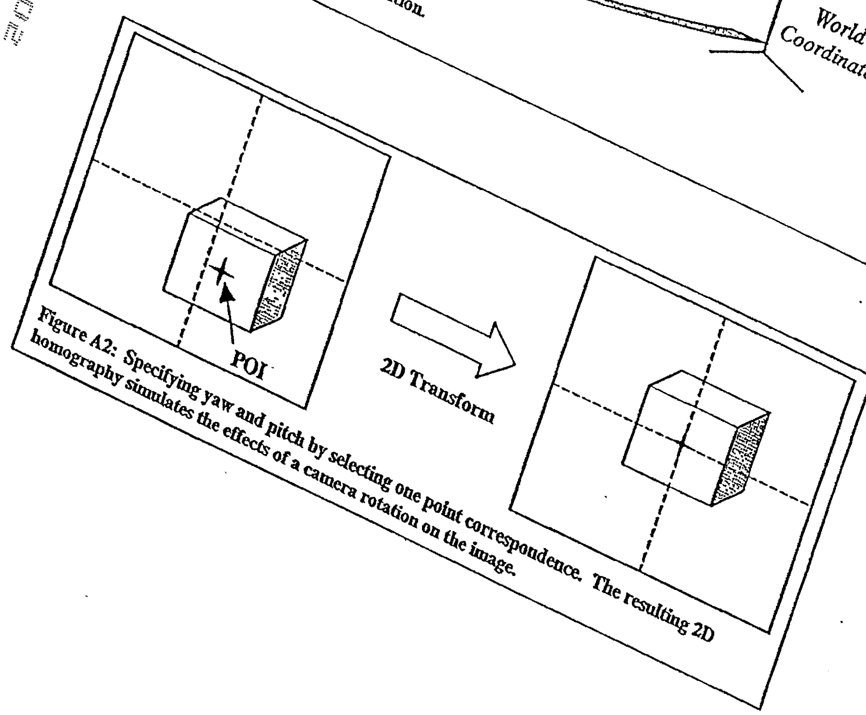

Figure A2: Specifying yaw and pitch by selecting one point correspondence. The resulting 2D homography simulates the effects of a camera rotation on the image.

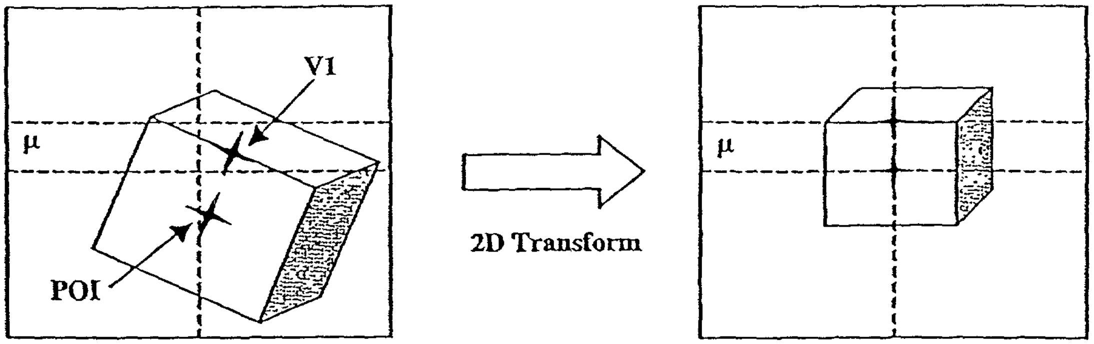

Figure A3: Specifying yaw, pitch, roll and Tz by selecting two point correspondences.

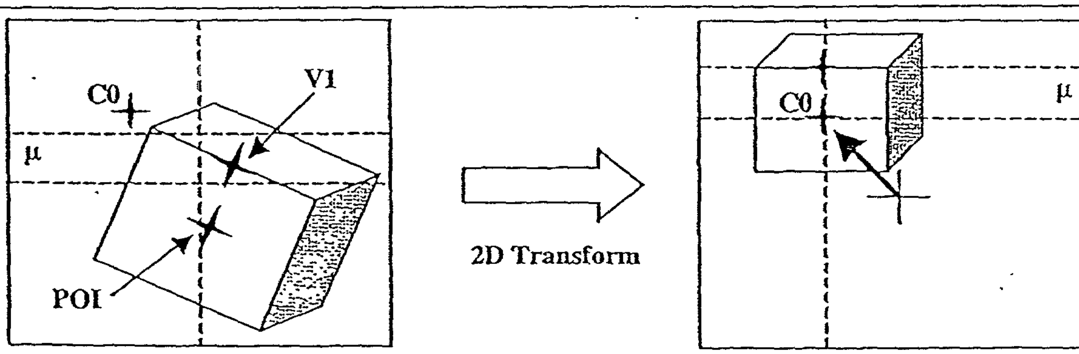

Figure A4: Specifying yaw, pitch, roll, Tx, Ty and Tz by selecting three points.

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division

Application deficiencies found during scanning:

☐ Page(s)___________ of ______________________________ were not present for scanning.
(Document title)

☐ Page(s)___________ of ______________________________ were not present for scanning.
(Document title)

☐ ***Scanned copy is best available.***

– There are 7 pages of Appendix including in specification.

– Drawing figures 4, 5 are very dark.